

PREVISIÓN DE LA DEMANDA DE BICICLETAS COMPARTIDAS MEDIANTE APRENDIZAJE AUTOMÁTICO: EL CASO DE VALLADOLID, ESPAÑA

JUAN S. ANGARITA-ZAPATA¹, FERRAN TORRENT-FONTBONA¹, BEATRIZ MACHÓ², FRANCISCO BENITO SÁNCHEZ³

¹ Aimsun, España.

²Autobuses Urbanos de Valladolid (AUVASA), España.

³Ayuntamiento de Valladolid, Agencia de Innovación y Desarrollo Económico de Valladolid, España.

PALABRAS CLAVE	RESUMEN
Bicicletas compartidas Previsión de la demanda Aprendizaje supervisado Aprendizaje no supervisado Aprendizaje automático	<i>Los sistemas de bicicletas compartidas funcionan mejor cuando los usuarios pueden contar con encontrar una bicicleta y un aparcamiento libre. Esta simple expectativa es difícil de cumplir durante las horas punta y los eventos especiales. Describimos un proceso práctico de previsión de la demanda a corto plazo creado mediante aprendizaje automático (ML) e implementado en Valladolid, España, como parte del proyecto SPINE financiado por la UE. Utilizando registros anónimos de una estación a otra del sistema de bicicletas compartidas de la ciudad, generamos previsiones de 60 minutos y elevamos esos pronósticos a los paneles de control de operaciones. Además, se evalúan estrategias de reequilibrio en un entorno de simulación. Resumimos el flujo de datos y la arquitectura del modelo, mostramos resultados representativos a nivel de red y describimos cómo el operador de bicicletas compartidas puede utilizar el sistema para resolver cuestiones diarias. Concluimos con todo lo aprendido y las prioridades a adoptar en los siguientes pasos.</i>

RECIBIDO: 05 / 09 / 2025

ACEPTADO: 20 / 10 / 2025

1. Introducción

Las bicicletas compartidas públicas se han convertido en una parte práctica de la movilidad cotidiana en muchas ciudades. Son baratas, eficientes en cuanto al espacio y limpias, y conectan a las personas con sus trabajos, escuelas y servicios sin aumentar el tráfico. Sin embargo, un punto débil es la disponibilidad de bicicletas. Los usuarios se encuentran con estaciones vacías cuando quieren iniciar un trayecto y con estaciones llenas cuando quieren terminarlo. Para un operador, adelantarse a estos desequilibrios es un reto diario, y depende de saber qué va a ocurrir en estaciones concretas en los próximos minutos, no mañana.

Ya existe una gran cantidad de investigaciones sobre la predicción de la demanda a corto plazo para el uso compartido de bicicletas. Los métodos van desde modelos de recuento sencillos hasta el aprendizaje automático basado en gráficos y el aprendizaje profundo que capturan tanto el espacio como el tiempo (Gammelli et al., 2022; Jiang, 2022; Li et al., 2023; Martín et al., 2023; Peláez-Rodríguez et al., 2024; Sohrabi et al., 2020; Yang et al., 2020). Algunos estudios incluso conectan las predicciones con las políticas de reequilibrio. Sin embargo, persisten dos lagunas prácticas. En primer lugar, muchos enfoques son difíciles de ejecutar como servicio de producción, asumen datos y condiciones estables que rara vez existen. En segundo lugar, tienden a limitarse a la precisión del modelo y dejan a los operadores sin las herramientas que necesitan: comprobaciones automáticas de datos, señales de supervisión sencillas y previsiones fáciles de leer y sobre las que actuar en tiempo real.

Este artículo aborda esas lagunas describiendo y evaluando un flujo de trabajo completo basado en el aprendizaje automático para la predicción a corto plazo a nivel de estación, desarrollado como parte del proyecto SPINE financiado por la UE. El flujo de trabajo consta de cuatro partes principales. En primer lugar, la limpieza de datos estandariza, anonimiza e imputa las llegadas y salidas a nivel de estación —FlowIN y FlowOUT— en intervalos de tiempo fijos. En segundo lugar, un módulo de patrones no supervisado agrupa los perfiles de demanda diarios típicos. En tercer lugar, un pronosticador supervisado que predice con 60 minutos de antelación por estación y adjunta indicadores de fiabilidad sencillos, de modo que el operador de bicicletas puede juzgar cuándo confiar en las cifras. Los resultados se introducen en un panel de control del operador y pueden utilizarse directamente en análisis hipotéticos para el reposicionamiento de bicicletas y estaciones.

En este contexto, las principales contribuciones de este trabajo son: (I) Detallamos un proceso práctico para la predicción a 60 minutos a nivel de estación —que incluye la limpieza de datos, el descubrimiento de patrones y la predicción de la demanda— que puede ejecutarse de forma continua con datos reales del operador. (II) Presentamos pruebas de una implementación en Valladolid, que muestran la alineación entre la dinámica de la red prevista y la observada, e ilustran los resultados típicos que son importantes para los operadores.

Una vez presentadas las principales contribuciones, este artículo se estructura de la siguiente manera. La sección 2 revisa los trabajos relacionados con la predicción de la demanda de bicicletas compartidas. La sección 3 describe los datos y el flujo de trabajo de predicción. La sección 4 presenta los resultados representativos. La sección 5 trata sobre la implementación en Valladolid y la interfaz del operador. La sección 6 analiza los siguientes pasos, incluida la integración con la simulación para la evaluación del reequilibrio basado en escenarios.

2. Previsión de la demanda de bicicletas: trabajos relacionados

A medida que los sistemas de bicicletas compartidas se convierten en una opción de transporte sostenible cada vez más frecuente en ciudades de diferentes tamaños, también ha aumentado el número de investigaciones sobre cómo gestionarlos y operarlos de una manera más eficiente. Como parte de este esfuerzo de investigación, el aprendizaje automático (ML) se ha convertido en un enfoque metodológico predominante para abordar el problema de la predicción de la demanda de bicicletas. Concretamente, el ML tiene la capacidad de modelar las complejas dependencias espaciotemporales de la demanda de bicicletas a nivel de estación. Así, es posible predecir con gran precisión la demanda esperada a diferentes niveles, por ejemplo, a nivel de estación.

Un ejemplo representativo de esta línea de investigación es el trabajo realizado por Sohrabi et al. Los autores desarrollaron modelos de recuento multivariante de valores extremos (GEV) para las llegadas y salidas a nivel de estación con términos de hora del día, clima y espacio. Este enfoque facilita la inferencia rápida y es estadísticamente interpretable. Sin embargo, tiene una capacidad limitada no lineal y una visión mayoritariamente local de la red que subestima los efectos indirectos a largo plazo y las complejas interacciones espaciales.

Partiendo de ahí, el estudio de Jiang destacó las contribuciones de los métodos de aprendizaje profundo (DL) (por ejemplo, CNN, RNN, GNN), que pueden codificar las interacciones entre estaciones, la dependencia temporal y el contexto exógeno. Un paso complementario hacia una estructura espacial más rica proviene de GNN a través del estudio de Yang et al. Los autores derivan descriptores de red con retraso temporal a partir de los flujos observados y los introducen en las líneas de base XGBoost/MLP/LSTM, lo que eleva de forma constante la precisión a corto plazo en la predicción de la demanda de bicicletas en Nueva York y Chicago. La ventaja es la interpretabilidad y la portabilidad; la desventaja es la dependencia de gráficos OD precalculados y características creadas manualmente que pueden ser escasas para estaciones recién inauguradas.

Gao y Chen (2022) trabajaron con el sistema de Seúl, los autores combinan datos cotidianos (clima, calidad del aire, tráfico, pandemias e información socioeconómica básica) para ver qué es lo que realmente impulsa el alquiler de bicicletas. Los autores probaron varios algoritmos conocidos y descubrieron que Random Forests y Support Vector Machines, ofrecen la mejor precisión. Una comprobación de importancia muestra que los indicadores meteorológicos, de contaminación y de COVID-19 son los que más peso tienen, mientras que el contexto socioeconómico añade pistas más útiles, lo que proporciona a los operadores una receta práctica para elaborar previsiones de demanda más sólidas.

Existen trabajos de evaluación comparativa más amplios, por ejemplo, la comparación de Peláez-Rodríguez et al. (2024) de modelos de ML y DL para series temporales de demanda de bicicletas con el fin de caracterizar el papel de las covariables exógenas. Sin embargo, dicho estudio se centra en la metodología y no articula cómo deben integrarse las predicciones en las operaciones diarias. En esta misma línea, Subramanian et al. compararon herramientas de previsión basadas en el aprendizaje automático, desde modelos estadísticos clásicos hasta redes neuronales modernas, utilizando datos históricos de bicicletas compartidas. Como conclusión principal, los autores descubrieron que las redes recurrentes ligeras (especialmente las GRU) tienden a superar a los enfoques tradicionales para las previsiones a corto plazo a nivel de estación.

Pasando de la predicción a las decisiones, Gammelli et al. (2022) combinan una RNN de Poisson variacional con una función de insatisfacción del usuario para elegir los inventarios iniciales nocturnos, lo que demuestra que una mayor precisión en las previsiones no se traduce automáticamente en mejores resultados operativos, a menos que las capas de predicción y decisión se diseñen conjuntamente. Qiao et al. (2021) diseñaron una red neuronal convolucional que aprende cómo cambia la demanda de bicicletas entre estaciones y a lo largo del tiempo. En lugar de tratar el espacio y el tiempo como cuadrículas fijas, el modelo se adapta al contexto (hora punta de la mañana, tardes lluviosas, eventos especiales), lo que le ayuda a detectar las señales locales adecuadas cuando son importantes. Con datos reales de bicicletas compartidas, esta configuración adaptativa predice de forma consistente la demanda a corto plazo con mayor precisión que los métodos estándar.

A continuación, el trabajo de Martín et al. (2023) introdujo un marco de modelización que combina modelos basados en datos y modelos de simulación para apoyar a los operadores de movilidad compartida en su proceso de toma de decisiones y despliegues. Inicialmente, la demanda del servicio se estima utilizando un modelo de predicción basado en datos, entrenado a partir de datos históricos de las operaciones del servicio. A continuación, las características de la oferta del servicio se determinan mediante un módulo de optimización que calcula el número óptimo de estaciones y su ubicación para una zona de despliegue determinada. La herramienta es flexible a la hora de diseñar y evaluar diferentes escenarios basados en hipótesis sobre las estrategias de despliegue que se deben tener en cuenta.

Por último, Li et al. (2023) utilizaron CNN fusionado con LSTM para capturar «vecinos semánticos» (áreas distantes con perfiles de uso similares), lo que mostró mejoras en varias ciudades, especialmente en torno a los picos; el precio es la mayor complejidad de servicio (aprendizaje de adyacencia + modelización de secuencias), con una menor transparencia y unos presupuestos de latencia/computación más ajustados.

En conjunto, todos los estudios mencionados anteriormente marcan un avance real en la investigación, desde modelos de recuento interpretables hasta características sensibles a gráficos y redes convolucionales, pero ninguno ofrece un flujo de trabajo completo y apto para la producción que las ciudades puedan ejecutar en tiempo real. Por lo tanto, las principales lagunas se encuentran en torno a la canalización de datos y la solidez operativa, más que en la modelización, lo que pone de relieve que la configuración actual evalúa los modelos a posteriori, sin un proceso integral para la ingestión continua, la validación automatizada y el reentrenamiento para respaldar las predicciones en tiempo real y las decisiones de los operadores. Esto último implica una arquitectura operativa que incluye la limpieza automatizada de datos, el descubrimiento y la supervisión de patrones (agrupación, detección de patrones de demanda) y las predicciones automáticas.

Igualmente importante es que la mayoría de los estudios omiten la interfaz de usuario que realmente necesitan los operadores públicos: paneles intuitivos basados en funciones que convierten las predicciones en aspectos procesables sin necesidad de ser expertos en aprendizaje automático ni tocar el código del modelo. Sin esa capa, las implementaciones tienden a estancarse en la fase piloto o a seguir dependiendo de especialistas técnicos, lo que limita el impacto operativo real.

3. Metodología

3.1. Alcance del caso de uso

Actualmente, es bien sabido que la mejora de los nuevos servicios de movilidad, como el uso compartido de bicicletas, puede reducir el atractivo del uso del coche para dar prioridad al transporte público y a los medios de transporte relacionados. Estas mejoras pueden ir desde la ampliación de la infraestructura física que aumenta la cobertura del sistema, hasta herramientas de toma de decisiones que proporcionan información para gestionar mejor los sistemas de uso compartido de bicicletas ya existentes. En este sentido, este caso tiene como objetivo mejorar la gestión del sistema de bicicletas compartidas de Valladolid, que se compone de 97 estaciones de bicicletas repartidas por toda la ciudad, como se muestra en la Figura 1. Este caso de uso forma parte del proyecto SPINE, financiado por la UE.

Figura 1 Estaciones BIKI

Fuente: Elaboración propia, 2025.

Dentro del proyecto SPINE, buscamos proporcionar una herramienta digital al caso piloto de Valladolid, que nos permita evaluar el rendimiento del sistema de bicicletas compartidas a la hora de responder a las diferentes demandas de los usuarios que solicitan su uso a través de las estaciones de bicicletas disponibles. En concreto, este caso busca predecir la demanda del sistema de bicicletas compartidas (utilizando el flujo de trabajo basado en inteligencia artificial introducido en la sección 3.3) para conocer el número de bicicletas que llegan y salen de cada estación.

3.2. Conjuntos de datos y configuración experimental

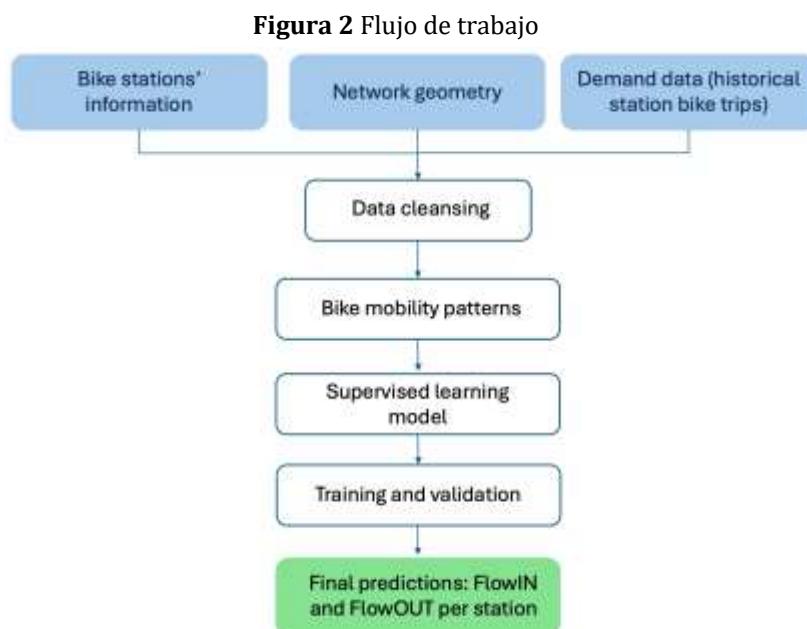
Se recopilaron los desplazamientos históricos en bicicleta de estación a estación del sistema de bicicletas compartidas de Valladolid. Estos registros corresponden a los desplazamientos en bicicleta realizados durante 2023 y 2024, que se anonimizaron para el desarrollo de la solución de aprendizaje automático. Los desplazamientos se componen principalmente de la hora de inicio y finalización, así como de los ID de las estaciones de inicio y finalización, a través de intervalos de tiempo irregulares. Por lo tanto, los desplazamientos se agregaron a intervalos de tiempo fijos (60 minutos) por estación. A continuación, se calculan dos nuevas variables para cada estación, que son FlowIN y FlowOUT. La primera representa las llegadas de bicicletas por hora, mientras que la segunda representa las salidas de bicicletas por hora, ambos casos a nivel de estación. Así, el conjunto de datos resultante se compone de una serie temporal, con 1.666.488 instancias, que muestra los valores de FlowIN y FlowOUT por estación y por marca de tiempo de 60 minutos para los dos años de datos disponibles.

En este sentido, abordamos la previsión de la demanda de bicicletas compartidas como un problema de regresión supervisada en el que las variables objetivo que se deben predecir son FlowIN y FlowOUT, a través de horizontes temporales de 60 minutos. Evaluamos la precisión de las predicciones utilizando la métrica del error cuadrático medio (RMSE). Además, el RMSE se calcula para la demanda total en todo el sistema BIKI y a nivel de estación, utilizando en ambos casos una validación cruzada quíntuple.

3.3. Flujo de trabajo para la predicción de la demanda de bicicletas compartidas

Esta solución digital implica un conjunto de procesos de datos, que funcionan conjuntamente para realizar predicciones de la demanda de bicicletas. Estos procesos se resumen en la Figura 2. Como

primer paso, las tres fuentes de datos principales son: la información relacionada con las estaciones de bicicletas (por ejemplo, coordenadas, capacidades de las estaciones), la geometría de la red que representa la infraestructura física por la que pueden circular las bicicletas y 3) los desplazamientos de estación a estación.



Fuente: Elaboración propia, 2025.

A continuación, el primer proceso es la limpieza de datos, cuyo objetivo es escanear y procesar los datos brutos de FlowIN y FlowOUT. Filtra los valores atípicos y poco realistas, e imputa los datos que faltan. Este paso consiste en tomar los datos brutos de entrada y llevar a cabo un muestreo ascendente o descendente para fragmentarlos y obtener una matriz de series temporales de FlowIN y FlowOUT (donde las columnas son estaciones y las filas son marcas de tiempo con las medidas de flujo correspondientes). En este sentido, se considera que la limpieza de datos garantiza la integridad del conjunto de datos y estandariza la frecuencia de muestreo de las estaciones. Además, también prepara un conjunto de datos adecuado para ser analizado y procesado en los siguientes pasos.

Una vez limpios los datos brutos, el siguiente paso es aplicar un algoritmo generador de patrones (PG). Este procesamiento se encarga de extraer los patrones diarios de movilidad en bicicleta a partir de los datos FlowIN y FlowOUT limpios. Estos patrones se identifican mediante técnicas de agrupamiento, un proceso de aprendizaje automático (ML) no supervisado diseñado para descubrir grupos de puntos de datos que comparten características similares. La idea principal es agrupar los días con perfiles diarios similares de FlowIN y FlowOUT, identificando de manera efectiva los patrones de tráfico comunes en períodos de 24 horas.

El algoritmo PG puede utilizar métodos como el agrupamiento jerárquico k-means. Estos métodos son adecuados porque permiten a los usuarios definir el número de grupos mediante la experimentación iterativa para encontrar la agrupación más adecuada. Además, se adaptan bien a medida que aumenta el número de muestras de datos. Una vez completada la agrupación, el siguiente paso es calcular el patrón representativo de cada clúster (es decir, grupo de días), teniendo en cuenta tanto la cantidad como la calidad de los datos.

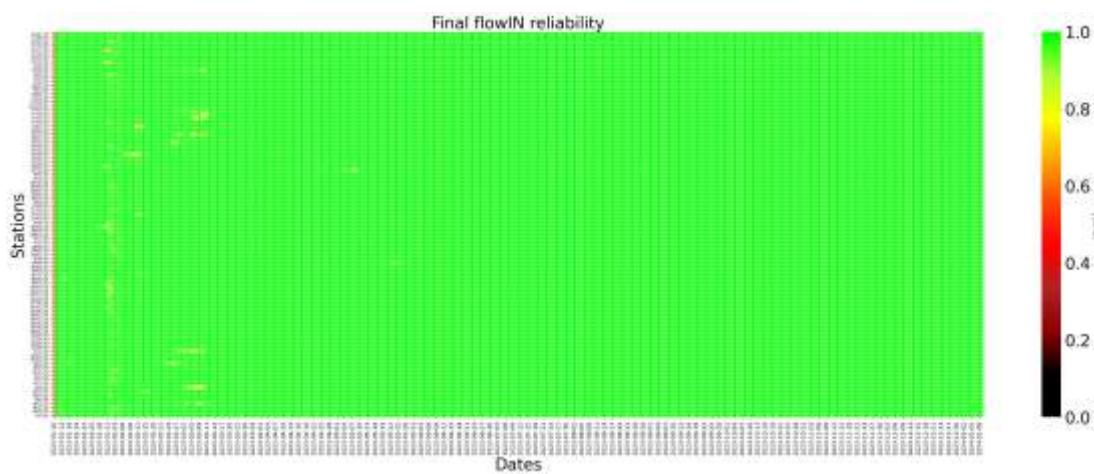
Tras identificar los patrones de tráfico representativos, el siguiente paso es realizar predicciones para estimar la demanda futura de bicicletas por estación. Para ello, los datos de entrada y la estación correspondiente deben asignarse al gráfico de red a intervalos de tiempo regulares (por ejemplo, cada 60 minutos). A continuación, se entrena y se prueba un modelo de aprendizaje automático (ML) supervisado mediante validación cruzada; en este caso, se utilizaron cinco pliegues. Este paso final lo realiza el módulo Network Mobility Forecaster (NMF), que puede proporcionar previsiones a corto y medio plazo, desde los próximos 15 a 60 minutos o incluso las

próximas horas, aprovechando datos históricos, en tiempo real. El NMF combina la agrupación para identificar tipos de días y un módulo aleatorio en su núcleo que se encarga de estimar la probabilidad de que el día objetivo sea uno de los disponibles. La información sobre el tráfico para cada tipo de día, junto con las probabilidades estimadas, se utiliza con observaciones recientes de las variables objetivo para predecir, por ejemplo, FlowIN y FlowOUT, como entrada para una capa convolucional gráfica que se encarga de proporcionar las predicciones a corto y medio plazo.

4. Resultados y análisis

Esta sección se centra en los principales resultados relacionados con los módulos presentados en la sección anterior. Figura 3 presenta una muestra de los mapas generados por el proceso de limpieza de datos, de las mediciones de FlowIN para las 97 estaciones del caso de Valladolid. El eje X representa el intervalo de tiempo de los datos disponibles, mientras que el eje Y indica la calidad de los datos, concretamente la fiabilidad o confianza de cada punto de datos, por día y por estación (cuanto más verdes son los valores, más fiables son los datos).

Figura 3 Datos limpios de FlowIN

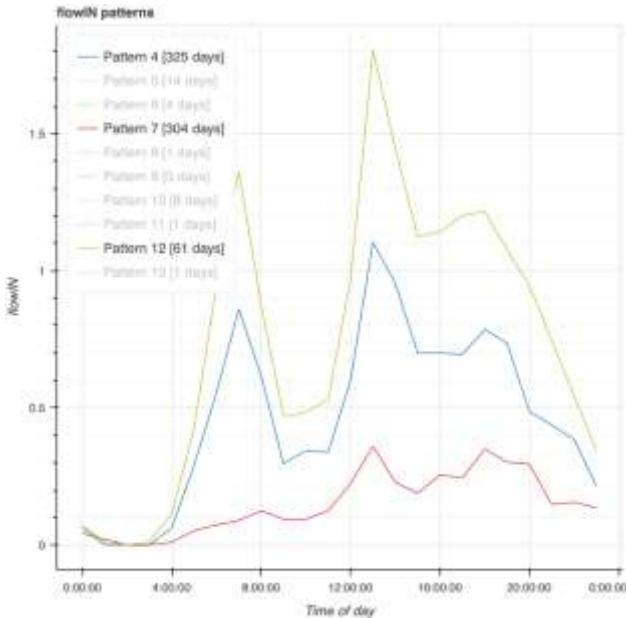


Fuente: Elaboración propia, 2025.

A continuación, la figura 4 muestra los patrones de FlowIN durante un periodo de 24 horas en el eje X. El número de días agrupados dentro de cada patrón también se indica en las etiquetas del gráfico. En general, los patrones capturan las fluctuaciones típicas del tráfico a lo largo del tiempo, como el bajo FlowIN a primera hora de la mañana, seguido de un aumento de la demanda durante el mediodía y la tarde. Concretamente, la

Figura 4 presenta tres patrones de FlowIN que resumen los comportamientos de movilidad en bicicleta más representativos durante los dos años de datos históricos. En concreto, los patrones 4 y 12 (series azul y dorada) ilustran los patrones de los días laborables, cuando hay una gran demanda de bicicletas (por ejemplo, de 11:00 a 15:00). A continuación, el segundo patrón representativo es el rojo, que representa el comportamiento de la demanda de bicicletas durante el fin de semana, a primera vista, menor, como era de esperar. Este patrón también incluye los días festivos, como el 1 de mayo y las vacaciones de Semana Santa.

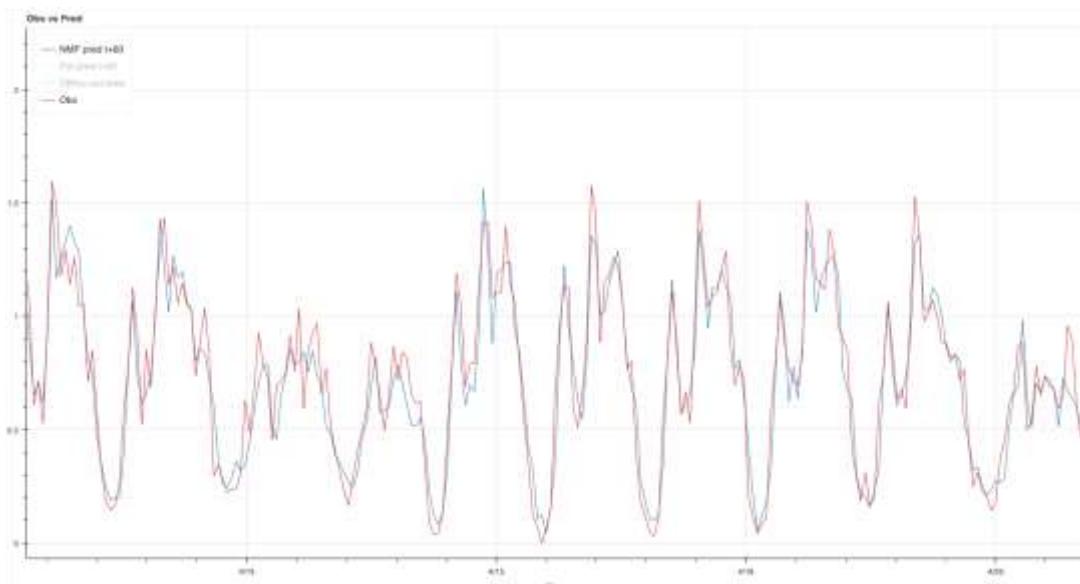
Figura 4 Principales patrones identificados por FlowIN



Fuente: Elaboración propia, 2025.

Figura 5 muestra la demanda media real y prevista de FlowIN agregada a través de las 97 estaciones a lo largo de 2023 y 2024. Como se puede observar, a nivel de red, las curvas de demanda previstas y observadas se alinean bien en las muestras de varios días, incluyendo el momento y la forma de los picos matutinos y vespertinos. Esto es precisamente lo que necesita el operador de bicicletas, es decir, saber cuándo y dónde se producirán los picos de demanda de bicicletas.

Figura 5. Demanda real y prevista de FlowIN

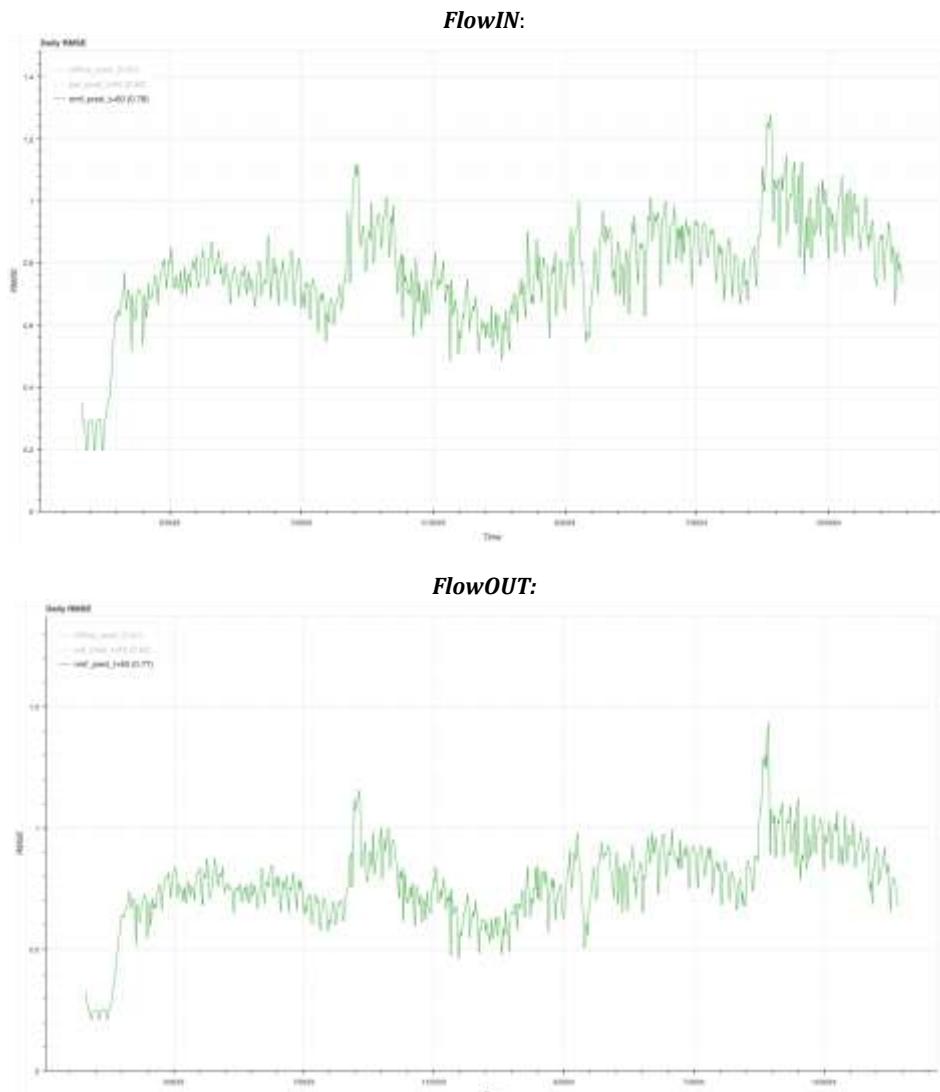


Fuente: Elaboración propia, 2025.

Una forma adecuada de corroborar la precisión de las predicciones es estimar el RMSE en cada observación de 60 minutos. Esto se muestra en la Figura 6, que resume los valores RMSE medios para todas las estaciones. Como se puede observar, el RMSE medio es de 0,78 y 0,77 para FlowIN y FlowOUT, respectivamente. Esto pone de relieve la calidad de la predicción, según la

interpretación de la métrica RMSE, que indica que cuanto más bajos son sus valores, mejor es el rendimiento del sistema evaluado.

Figura 6 RMSE para FlowIN y FlowOUT



Fuente: Elaboración propia, 2025.

5. Implementación en Valladolid

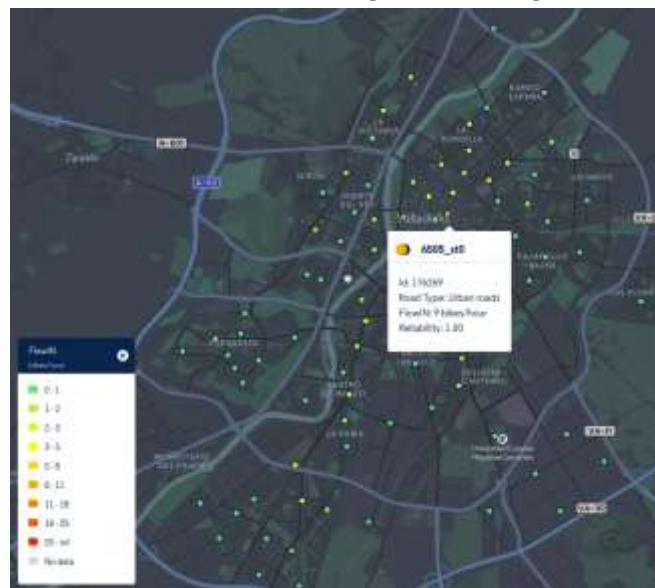
El flujo de trabajo y los resultados que se presentan a continuación se han implementado en un panel de control como parte de la solución Aimsun Predict¹. Dentro del proyecto SPINE, la idea principal detrás de esta implementación es que las partes interesadas del proyecto piloto puedan acceder al panel de control y visualizar el patrón y las predicciones, sin necesidad de interactuar directamente con los modelos de aprendizaje automático. En su lugar, tienen acceso a una interfaz fácil de usar que emula el tiempo real, como se muestra en la Figura 7.

El panel de control presenta un mapa de la ciudad con un fondo oscuro salpicado de pequeños marcadores circulares. Cada punto representa una estación de bicicletas. Los puntos están codificados por colores para mostrar FlowIN o FlowOUT. Cuando el usuario hace clic en una estación, se abre una ventana emergente blanca que muestra el nombre y los detalles de la estación, por ejemplo: «Id: 6505_st0, Tipo de carretera: Carreteras urbanas, FlowIN: 9

¹ <https://www.aimsun.com/aimsun-predict>

bicicletas/hora, Fiabilidad: 1,00». El punto de esa estación tiene un color que coincide con el rango de 9 bicicletas/hora, dentro del panel izquierdo, donde se resaltan todos los rangos de demanda de FlowIN. En concreto, este panel distingue los rangos de demanda con colores, los mismos que se utilizan para los puntos de las estaciones en el mapa.

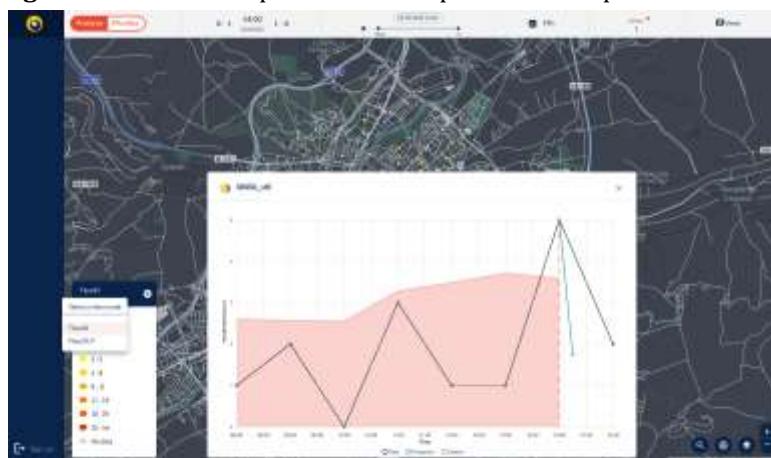
Figura 7 Estaciones de bicicletas disponibles en el panel de control



Fuente: Elaboración propia, 2025.

El usuario también puede abrir los detalles de cada estación (véase**Figura 8**). En ella, es posible ver la demanda histórica (ya sea FlowIN o FlowOUT, dependiendo de la variable seleccionada en el panel izquierdo) de la estación en cuestión. Además, para cada estación también es posible ver el patrón al que corresponde el día revisado y la predicción hasta la siguiente hora con antelación.

Figura 8 Predicciones por estación disponibles en el panel de control



Fuente: Elaboración propia, 2025.

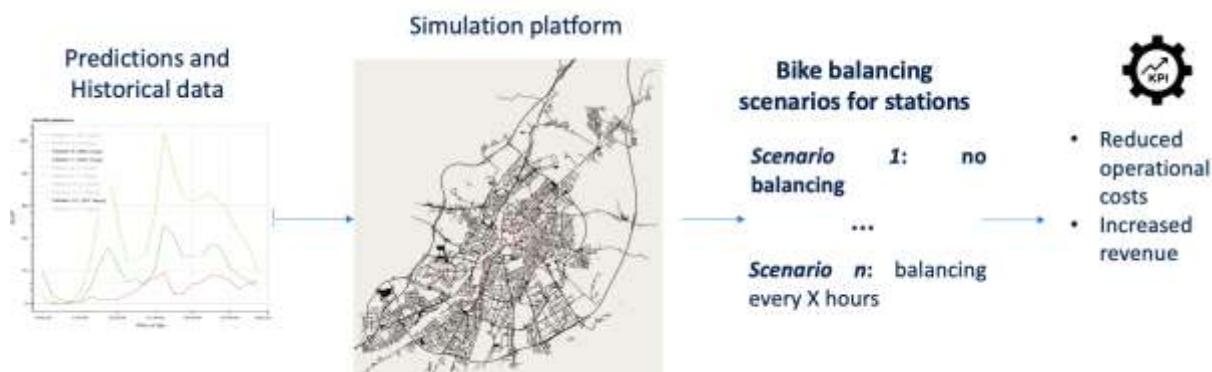
También es importante mencionar que el flujo de trabajo integral presentado en este documento puede replicarse en otras ciudades con sus propios sistemas de bicicletas compartidas, que pueden variar en cuanto a cobertura espacial (por ejemplo, centros urbanos, áreas metropolitanas), tipos de bicicletas (por ejemplo, mecánicas, eléctricas), entre otras características. Esto último es posible porque los componentes algorítmicos de Aimsun Predict pueden transferirse a otras ciudades, siempre que se garanticen los siguientes requisitos mínimos de datos: 1) Datos históricos y/o en tiempo real de la demanda de desplazamientos en bicicleta

(por ejemplo, viajes entre estaciones y ocupación de las mismas, etc.); 2) Mapa abierto de la ciudad y/o zona de interés y 3) Transferencia de datos a través de API u otros medios para el intercambio de datos entre Aimsun y el departamento de TI del operador de transporte público de la ciudad.

6. Próximos pasos

Una vez finalizado el flujo de trabajo de ML y presentados los resultados, el siguiente paso dentro del proyecto SPINE consistirá en utilizar la simulación para evaluar diferentes estrategias de reequilibrio de la flota a través de múltiples horizontes temporales; por ejemplo, reequilibrando cada 1, 3 o 4 horas, como se muestra en la Figura 9. Esto permitirá mejorar la capacidad de planificación del sistema de bicicletas compartidas, proporcionando información sobre el diseño del servicio que prioriza la disponibilidad de bicicletas en las estaciones más estratégicas y de mayor demanda de la ciudad. En conjunto, estas herramientas forman un Gemelo Digital de Movilidad Compartida (Shared Mobility Digital Twin), un potente marco para la gestión proactiva y optimizada de la flota de bicicletas. El resultado: mayor satisfacción de los usuarios, reducción de los costes operativos y aumento de los ingresos para el operador de bicicletas compartidas.

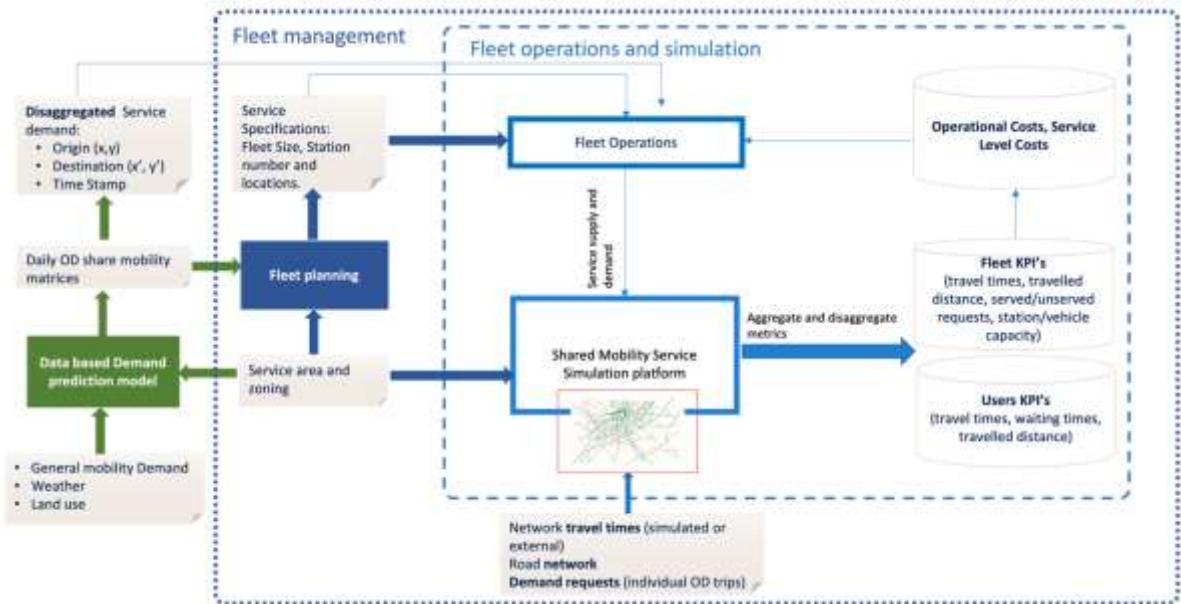
Figura 9 Marco de simulación conectado a las predicciones para el proyecto piloto de Valladolid



Fuente: Elaboración propia, 2025.

Como ejemplo para ilustrar cómo funcionará esta herramienta de simulación, podemos hacer referencia al trabajo realizado en el proyecto MOMENTUM, financiado por la UE (Martín et al., 2023). Dicho marco combinaba enfoques de modelización y simulación para evaluar y comparar diferentes estrategias de despliegue del servicio. Se compone de tres pasos que se describen a continuación y se resumen en la Figura 10.

Figura 10 Flujo de trabajo de MOMENTUM (Martín et al., 2023)



Fuente: Elaboración propia, 2025.

- **Predicción de la demanda:** se estima la demanda de un servicio de movilidad compartida en una zona determinada. Para dicha estimación, se ha desarrollado un modelo de aprendizaje automático basado en la regresión. El modelo predictivo se entrenó con la demanda histórica del servicio en aquellas zonas de la ciudad en las que el servicio ya está implantado y se aplicó a las zonas en las que los autores querían explorar las estrategias de implantación. El resultado de este modelo fue una matriz OD para las áreas de interés. Esta se utilizó para: i) identificar la distribución óptima de las estaciones; y ii) usar el modelo de simulación para medir el rendimiento y el impacto de la prestación del servicio en el área de estudio, para lo cual se disagregó la matriz OD obtenida con el fin de proporcionar posiciones temporales y espaciales específicas.
- **Planificación de la flota:** en esta parte, se ejecutó un algoritmo de optimización para determinar las características del servicio, como el número de estaciones, el tamaño y las características de la flota, o la ubicación óptima para cubrir la demanda estimada. El algoritmo apoyó la provisión de soluciones sujetas a restricciones relacionadas con el número máximo y mínimo de estaciones a considerar, así como la distancia máxima a pie desde el origen hasta el destino final de la estación.
- **Operaciones y simulación de la flota:** por último, las características estratégicas del servicio y las estimaciones de la demanda derivadas de los pasos anteriores se utilizaron como entrada en la plataforma de simulación para servicios de movilidad compartida, que está integrada con el algoritmo operativo para optimizar y simular las operaciones del servicio durante un período de tiempo. En particular, el operador tuvo en cuenta la demanda del servicio en cada momento en que se realiza una solicitud, la disponibilidad de vehículos, así como la información relativa a los tiempos y distancias de desplazamiento de la red que se obtienen de la simulación. El resultado de este paso proporcionó una lista de indicadores que permiten evaluar el rendimiento del servicio tanto a nivel de usuario como de sistema.

7. Agradecimientos

Este trabajo ha sido financiado por el Programa de Investigación e Innovación Horizonte Europa de la Unión Europea a través del proyecto «INICIATIVAS DE TRANSPORTE PÚBLICO INTELIGENTE PARA CIUDADES CLIMÁTICAMENTE NEUTRAS EN EUROPA», con el número de acuerdo de subvención 101096664.

Referencias

- Gammelli, D., Wang, Y., Prak, D., Rodrigues, F., Minner, S., & Pereira, F. C. (2022). Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management. *Transportation Research Part C: Emerging Technologies*, 138, 103571. <https://doi.org/10.1016/j.trc.2022.103571>
- Gao, C., & Chen, Y. (2022). Using machine learning methods to predict demand for bike sharing. In J. L. Stienmetz, B. Ferrer-Rosell, & D. Massimo (Eds.), *Information and Communication Technologies in Tourism 2022* (pp. 282-296). Springer. https://doi.org/10.1007/978-3-030-94751-4_25
- Jiang, W. (2022). Bike sharing usage prediction with deep learning: A survey. *Neural Computing and Applications*, 34, 15369–15385. <https://doi.org/10.1007/s00521-022-07380-5>
- Li, X., Xu, Y., Zhang, X., Shi, W., Yue, Y., & Li, Q. (2023). Improving short-term bike sharing demand forecast through an irregular convolutional neural network. *Transportation Research Part C: Emerging Technologies*, 147, 103984. <https://doi.org/10.1016/j.trc.2022.103984>.
- Martín, I., Tympakanaki, A., Cantú-Ros, O. G., Burrieza-Galán, J., Casas, J., Narayanan, S., Ayfantopoulou, G., Antoniou, C., & Maleas, Z. (2023). Data-driven modelling and simulation evaluation of shared mobility services: A case study for bike-sharing in Madrid. *102nd Annual Meeting of the Transportation Research Board*. Washington, DC.
- Peláez-Rodríguez, C., Pérez-Aracil, J., Fister, D., Torres-López, R., & Salcedo-Sanz, S. (2024). Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches. *Expert Systems with Applications*, 238, 122264. <https://doi.org/10.1016/j.eswa.2023.122264>
- Qiao, S., Han, N., Huang, J., Yue, K., Mao, R., Shu, H., He, Q., & Wu, X. (2021). A dynamic convolutional neural network based shared-bike demand forecasting model. *ACM Transactions on Intelligent Systems and Technology*, 12(6), Article 70. <https://doi.org/10.1145/3447988>
- Subramanian, M., Cho, J., Veerappampalayam Easwaramoorthy, S., Murugesan, A., & Chinnasamy, R. (2023). Enhancing sustainable transportation: AI-driven bike demand forecasting in smart cities. *Sustainability*, 15(18), 13840. <https://doi.org/10.3390/su151813840>
- Sohrabi, S., Paleti, R., Balan, L., & Cetin, M. (2020). Real-time prediction of public bike sharing system demand using generalized extreme value count model. *Transportation Research Part A: Policy and Practice*, 133, 325–336. <https://doi.org/10.1016/j.tra.2020.02.001>
- Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2020). Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. *Computers, Environment and Urban Systems*, 83, 101521. <https://doi.org/10.1016/j.compenvurbsys.2020.101521>